PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G10L 9/06

A1

(11) International Publication Number: **WO 99/48088**

(43) International Publication Date: 23 September 1999 (23.09.99)

(21) International Application Number: PCT/US99/06072

(22) International Filing Date: 19 March 1999 (19.03.99)

(30) Priority Data:
60/078,937 20 March 1998 (20.03.98) US

(71) Applicant *(for all designated States except US)*: INROAD, INC. [US/US]; Market Place Tower PH–B, 2025 First Avenue, Seattle, WA 98121 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: PROFIT, Jack, H., Jr. [US/US]; 14746 Glenacres Road, S.W., Vashon Island, WA 98070 (US). BROWN, N., Gregg [US/US]; 2011 Fifth Avenue, N., Seattle, WA 98109 (US).

(74) Agent: INOUYE, Patrick, J., S.; Christensen O'Connor Johnson & Kindness PLLC, Suite 2800, 1420 Fifth Avenue, Seattle, WA 98101 (US).

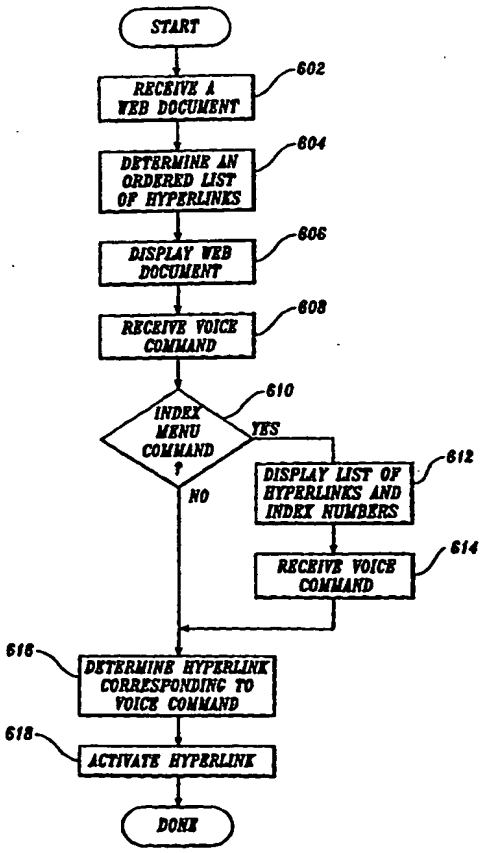
(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: VOICE CONTROLLED WEB BROWSER

(57) Abstract

A system and method for implementing a voice–controlled Web browser program executing on a wearable computer is disclosed. A Web document is received (602) at the wearable computer, and processed to dynamically generate a speech grammar (604). The speech grammar is used to recognize voice commands (616) at the wearable computer. Alternatively, a Web document is precompiled at a server computer to generate a speech grammar, and the speech grammar is transmitted with its corresponding Web document to the wearable computer. The wearable computer provides three mechanisms for a user to navigate Web pages by the use of voice. In one mechanism, an index value corresponding to each hyperlink is appended to the hyperlink text and displayed to the user (612). The user may speak the index value to activate the corresponding hyperlink (614). In a second mechanism, the user can speak the test of the hyperlink to activate the hyperlink (616). In a third mechanism, the user invokes a command to display a dialog window having a list of hyperlinks and their corresponding index values. The user can speak an index value or a hyperlink to activate the hyperlink (618).



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

VOICE CONTROLLED WEB BROWSER

This application claims the benefit of U.S. Provisional Application No. 60/078,937, filed March 20, 1998.

Field of the Invention

The present invention relates to the field of Web browsers and, in particular, to methods and systems for controlling a Web browser by the use of voice commands.

Background of the Invention

In recent years, there has been a tremendous proliferation of computers connected to a global network known as the Internet. "Client" computers download and upload digital information from "server" computers via the Internet. Client application software executing on such client computers typically accept commands from users and obtain data and services from the server computers by sending requests to server applications running on the server computers. Intranets are local area network containing one or more Web servers and client computers operating in a manner similar to the Internet as described above. Typically, all of the computers interconnected via an intranet operate within a company or organization.

A number of specialized protocols are used for exchanging commands and data between computers interconnected via the Internet, as are well known in the art.

The protocols include the file transfer protocol (FTP) used for exchanging files and the hypertext transfer protocol (HTTP) used for accessing data on the World Wide Web, often referred to simply as "the Web."

The Web is an information service on the Internet providing documents and hyperlinks between documents. The Web is made up of numerous Web sites around the world that maintain and distribute electronic documents. A Web site may use one or more Web server computers that store and distribute documents in various formats, including the hypertext markup language (HTML).

An HTML document contains text and metadata, that is, commands providing formatting information. HTML documents also include embedded "hyperlinks" that reference other data or documents located on any Web server computer. The referenced documents may represent text, graphics, audio, or video in respective formats.

A Web browser is a client application or operating system utility program that communicates with server computers via one or more Internet protocols, such as FTP and HTTP. Basically, Web browsers receive electronic HTML documents from server computers over the network and present them to users. The HotJava Web browser, available from Sun Microsystems, Palo Alto, California, is an example of a popular Web browser application.

There are many jobs in manufacturing and other industries where workers require access to information available through a computer terminal, but must also move around and work with their hands. The information includes company data residing on a network server, company Intranet information, or even information available on the Internet. These workers have been characterized as "locally mobile" workers.

By way of example, a locally mobile production worker might need access to blueprints, reference manuals, and the like, to properly perform a particular job. At present, to retrieve such information, this worker would have to cease working, leave their workspace, obtain the information, and return to their workspace. Some information may not be transportable. Even if the worker could return with the necessary information, the demands of the job may still make it extremely difficult for the worker to view the retrieved information while performing manual tasks at the same time.

By way of further example, many inventory-based jobs require the creation and physical dissemination of tremendous amounts of paperwork such as in, for

example, a distribution center, a worker may fill out invoice documents, and then send copies of the documents to the shipping department, the accounting department, the production department, and others. Such jobs have the potential for errors in the initial creation of such information. Additionally, the created information can be lost in the subsequent dissemination process. Time delay in disseminating such information may pose an additional problem.

Therefore, it would be desirable to provide a system and method for enabling a locally mobile worker with access to information needed to perform a job, without a worker having to leave the workspace. There is also a need for providing the locally mobile worker with a way to review the information while simultaneously performing the job. Preferably, the system and method would transmit data between a server computer and a mobile computer carried by the mobile worker.

It would be further desirable to have a system and method wherein a user employs a browser program to view and enter information, and wherein voice commands are used to control the browser program. Preferably, such a system and method would include a mechanism that allows a user to navigate between information pages and also allows a user to manipulate user interface controls by the use of voice commands.

Preferably, such a system and method would include alternate mechanisms for creating a speech-recognition grammar. One desirable mechanism includes the dynamic creation of a speech-recognition grammar after an information page is received by the user. A second desirable mechanism includes precompiling the information page to create a speech-recognition grammar that is transmitted with the information page to the user's computer. The present invention is directed to providing such a system and method with such associated mechanisms.

Summary of the Invention

The present invention includes a voice-activated Web browser program executing on a wearable computer. The browser program provides three mechanisms for allowing a user to employ voice commands to navigate pages. In one mechanism, a "speech hint," or index value corresponding to each hyperlink in a Web page is determined and displayed on the Web page. Preferably, a unique identifier, or index value, is appended to the end of the hyperlink text. When a voice command is received, a determination is made of whether the voice command corresponds to the index value. If the voice command corresponds to an index value, the hyperlink corresponding to the index value is activated to retrieve additional data.

In the second mechanism, when a voice command is received, a determination is made of whether the voice command corresponds to the text associated with a hyperlink on the current Web page. If the voice command corresponds to the text associated with a hyperlink, the associated hyperlink is activated to retrieve additional data.

In the third mechanism, a voice command causes a list of hyperlinks to be displayed. Each hyperlink is displayed with a corresponding index value. In response to receiving a voice command, a determination is made of whether the voice command corresponds to either hyperlink text or an index value corresponding to a hyperlink. If a match is found, the corresponding hyperlink is activated to retrieve additional data. Preferably, all three mechanisms are presented to a user, providing a user with a choice of using any mechanism to control the browser program.

In an additional aspect of the invention, an external speech grammar referenced by the Web document is dynamically compiled by the Web browser after receiving a Web document. The speech grammar is activated by the Web browser for use in processing subsequent voice commands whenever the Web document in question is displayed. This mechanism allows Web document developers to customize the speech features of a specific Web page.

In another aspect of the invention, a speech grammar corresponding to a Web document is compiled on a server computer and stored at the server computer. When a Web document is transmitted from the server computer to the Web browser, the corresponding compiled speech grammar is transmitted to the Web browser. The speech grammar is received at the browser and used to process voice commands pertaining to the Web page. This mechanism is similar to the one described in the previous paragraph. The main difference is that this mechanism takes advantage of the high performance of the server machine in compiling specific grammars.

The present invention provides a mechanism for controlling a browser program executing on a wearable computer by the use of voice commands. By providing five different mechanisms, the invention provides flexibility to a user. The mechanism also provides flexibility to a Web page author, who may optimally design the Web page to be used according to one or more of the mechanisms. The invention also provides a mechanism for controlling a browser when a Web page author has not designed the Web page which might include voice-activated control.

By providing a mechanism for pre-compiling speech grammars and a mechanism for dynamically compiling speech grammars, the invention can be used

when a Web page author has built a Web page with a speech grammar or when a Web page author has not built a speech grammar corresponding to the Web page.

Brief Description of the Drawings

The foregoing aspects and many of the attendant advantages of this invention will become more readily appreciated and better understood by reference to the following detailed description, when taken in conjunction with the accompanying drawings, wherein:

FIGURE 1A is a block diagram of a wearable computer system for implementing the present invention;

FIGURE 1B is a pictorial illustration of the wearable computer system of FIGURE 1;

FIGURE 2 illustrates an exemplary Web page displayed on a wearable computer, in accordance with the present invention;

FIGURE 3 is a block diagram illustrating a system for implementing a voice-controlled Web browser in accordance with the present invention;

FIGURE 4 is a block diagram illustrating an alternative system for implementing a voice-controlled Web browser;

FIGURE 5 is a flow diagram illustrating a process of generating a speech-recognition grammar for use in a voice-controlled Web browser program; and

FIGURE 6 is a flow diagram illustrating the process of displaying a Web document and handling a voice command, in accordance with the present invention.

Detailed Description of the Preferred Embodiment

The present invention is a mechanism and method for implementing a voice-controlled Web browser program executing on a wearable computer that communicates with one or more server computers. The mechanism and method of the invention generate a voice-recognition grammar. Upon receipt of a voice command, the mechanism and method of the invention utilize the voice-recognition grammar to determine which command was received and the received command is used to control and manipulate the Web browser program.

In accordance with the present invention, a Web browser program executes on a wearable computer. FIGURE 1A and the following discussion are intended to provide a brief, general description of a wearable computer upon which the invention may be implemented.

With reference to FIGURE 1A, an exemplary system for implementing the invention includes a wearable computer 102, including a central processing unit 104

(CPU), a system memory 106, and a system bus 108 that couples various system components, including the system memory 106, to the processing unit 104. The system memory 106 may include both volatile and nonvolatile memory (not shown).

A second bus, such as a PCI bus 110, communicates with the system bus 108 and transfers data to and from peripheral components. A video controller 112 connected to the PCI bus 110 controls the display of information on a video screen 114. An audio controller 116 connected to the PCI bus 110 controls a speaker device 118. The speaker device 118 may optionally be built into a headset 134 (shown in FIGURE 1B). The audio controller 116 also receives inputs from a microphone 120.

The wearable computer 102 includes various other components, such as a power supply and a system clock, that are not illustrated in FIGURE 1. A wearable computer system for use with the present invention is described in commonly assigned U.S. Patent Application, Serial No. 09/045,260, pending, the disclosure of which is incorporated herein by reference in its entirety.

FIGURE 1B illustrates an embodiment of a wearable computer 102 that is used to implement the present invention. A CPU 104 and a memory 106 (shown in FIGURE 1A) are contained within a base unit 130 that may be attached to a belt 132. A headset 134 includes a speaker device 118, a display screen 114, and a microphone 120.

The wearable computer 102 communicates with a server computer (not shown). The server computer transmits Web documents, such as HTML documents, to the wearable computer 102, which displays the documents to a user. In one embodiment, the documents are displayed on the display screen 114. The wearable computer may also play audio data via the speaker device 118. In an alternate embodiment, the video screen 114 is not present or is inactive. The video screen 114 may also be employed to selectively present Web documents, while other select Web documents are played only as audio data via the speaker device 118.

In response to the presentation of a Web document, a user may control the presentation of additional data and may transmit data to the server computer using voice commands. FIGURE 2 illustrates an exemplary Web document 150 that is displayed on the video screen 114. The Web document 150 contains hyperlinks, which each include a representative symbol, such as text or a graphic symbol. The symbol may also be an audio signal that is presented to the user. The representative symbol is referred to as an "anchor tag" 152. Each hyperlink also includes an

embedded address (not shown) corresponding to additional data. When a user selects a hyperlink, the additional data corresponding to the associated address is retrieved and presented to the user. A Uniform Resource Locator (URL) is one form of addressing that is commonly used in Web documents. An address can be a file system pathname or other value used to indicate the location of additional data.

The present invention includes three mechanisms that allow a user to employ voice commands to navigate Web pages: speakable indices, an index menu, and "speakable hyperlinks." Preferably, all three mechanisms are included, and a user has the option of using one or more of the mechanisms.

The speakable indices mechanism includes a speech-specific parser 206 (shown in FIGURES 3 and 4) that dynamically inserts a visual speech hint 154 next to each HTML anchor tag 152. Preferably, the speech hint 154 is an index number and is inserted immediately before each corresponding anchor tag 152. The superscripted index number is incremented with each successive anchor tag 152, so duplicate index numbers do not occur. As illustrated in FIGURE 2, the speech hint 154 appears before each hyperlink anchor tag 152.

The speakable indexing feature is preferably enabled and disabled via two speakable commands: "index enable" and "index disable." When a user speaks the words "index enable," the speakable index feature is enabled. When a user speaks the words "index disable," the speakable index feature is disabled. When enabled, this feature allows a user to speak a hyperlink tag's unique index number to follow the hyperlink. When an index number is spoken, a speech-recognition engine 212 (shown in FIGURE 3) generates a corresponding speech event, which is translated into a user command to follow the corresponding hyperlink.

The index menu mechanism provides a second method of following hyperlinks. When a user speaks the phrase "index menu," the mechanism and method of the invention displays a dialog box to the user. The dialog box includes a scrollable list of hyperlinks and their associated unique indices. A user may navigate this list using verbal scrolling commands, or may speak the unique index number corresponding to the hyperlink that they wish to follow.

To employ the speakable hyperlinks mechanism, a user speaks the contents of a hyperlink anchor tag 152. In response, the speech-recognition engine 212 (shown in FIGURE 3) generates a corresponding speech event that is translated into a user command to follow the corresponding hyperlink. An HTML rendering engine 208 navigates to linked Web content based on the user selection. Preferably, a Web page

author anticipating the use of speakable hyperlinks creates a Web page that does not have two hyperlink anchor tags that may sound similar.

Controls and images also have corresponding speech hints. For example, as depicted in FIGURE 2, selection controls 156 have corresponding speech hints 158. Edit controls 160 have corresponding speech hints 162. The image 164 has a corresponding speech hint 166 positioned at the upper left corner. Activating a control sets the focus of the browser to the control, so that additional voice input is directed to the control. The use of speech grammars to select controls is similar to the use of speech grammars to select hyperlinks.

Thus, the present invention provides a mechanism and method for dynamically generating speech grammars upon receipt of Web pages, and a mechanism and method for pre-compiling speech grammars prior to transmitting Web pages to the wearable computer. FIGURE 3 is a functional block diagram which illustrates components of a wearable computer system 200 that dynamically generates speech grammars upon receipt of Web pages. An HTML parser 204 receives an HTML document from the Internet or an intranet 202 and parses the document content to generate an internal representation 205 of the HTML document. The internal representation 205 is passed to a speech-specific parser 206. The speech-specific parser 206 locates hyperlinks or other interactive controls that may be the target of a voice command and generates speech grammars 209. The speech-specific parser 206 also generates visual speech hints 154 (shown in FIGURE 2). The revised internal representation 207 of the HTML document 205, with the visual speech hints generated by the speech-specific parser 206, is passed to an HTML rendering engine 208. The HTML rendering engine 208 generates a visual Web page 150 (shown in FIGURE 2) based upon the revised internal representation of the HTML document 207. The visual Web page 150 is displayed on the display screen 114 (shown in FIGURE 1A).

Speech grammars 209 are generated from the HTML text by the speech-specific parser 206 and are passed to a speech grammar compiler 210. The speech grammar compiler 210 translates the speech grammars 209 into a compiled speech grammar 211 that is used by a grammar-based speech-recognition engine 212. Many speech engine providers, including IBM and Lernout & Hauspie, provide grammar compilers with their speech engine products. In addition to the compiled speech grammars, the speech-recognition engine 212 receives static, or precompiled, grammars 214 that are used for controlling the Web browser. The static

grammars 214 include browser commands that are not Web page specific, such as "back" and "forward." The speech-recognition engine 212 receives voice audio input from the microphone 120 and uses the compiled speech grammars 211 and static speech grammars 214 to determine the command or text spoken into the microphone 120. Via Voice, a product licensed by IBM, is a commercially available speech-recognition engine that can be used as the speech-recognition engine 212 in the present invention.

In response to voice audio input, the speech-recognition engine 212 generates speech events 213. The speech events 213 generated by the speech-recognition engine 212 are handled by corresponding software speech controls 218. The speech controls 218 translate the speech events 213 into user commands 215, which are passed to the HTML rendering engine 208. In response to the receipt of user commands 215, the HTML rendering engine 208 performs an action corresponding to the user commands 215. For example, if a user command 215 designates that a particular hyperlink has been selected by a voice audio input, the HTML rendering engine 208 performs the action of retrieving the Web page corresponding to the hyperlink. Similarly, graphical user interface (GUI) controls 216, such as buttons or menus, can receive input from a mechanical device, such as a mouse, or other control (not shown). The GUI controls 216 generate user commands 217, which are passed to the HTML rendering engine 208 for appropriate handling, as described above. The speech controls 218 may also generate audio prompts 218, which are presented to the user via a headset or other speaker device 118.

FIGURE 4 is a functional block diagram which illustrates a voice-controlled Web browser system 300 that uses precompiled speech-recognition grammars 214. The system 300 is similar to the system 200 illustrated in FIGURE 3. The following discussion describes the important differences between the system 200 which uses dynamically generated speech-recognition grammars 209 and the system 300 which uses precompiled speech-recognition grammars 214.

In the precompiled system 300, a speech grammar compiler, such as the speech grammar compiler 210 illustrated in FIGURE 3, is used to generate a voice-recognition grammar at a Web server operating over the Internet or an intranet 202. A speech-specific parser 206 on the wearable computer receives an internal representation of the HTML document 205 and one or more previously compiled speech grammars 302. The speech-specific parser 206 passes the received speech grammars 213 to the grammar-based speech-recognition engine 212. The speech-

recognition engine 212 receives compiled speech grammars similar to the dynamic grammar system 200 of FIGURE 3. The precompiled grammar system 300 need not include the speech grammar compiler 210 of FIGURE 3.

As with the dynamic grammar system 200 of FIGURE 3, the speech-specific parser 206 passes the revised internal representation of the HTML document 207 to the HTML rendering engine 208. The HTML rendering engine 208, the GUI controls 216, the speech controls 214, and the speech-recognition engine 212 perform operations as described above with respect to the dynamic grammar system 202 of FIGURE 3.

In one embodiment of the invention, commands pertaining to speech grammars are embedded within HTML comment fields. The following HTML code segment shows, by way of example, instructions used to specify the location of a dictionary and grammar files.

```
<!–InroadSpeechGrammar(%text)–>
<!–InroadSpeechGrammar
        name = %NameOfElement
        grammar = %URL
        dictionary = % URL
–>
```

%NameOfElement specifies the name of an HTML element and corresponds to the name in an HTML hyperlink anchor tag 152. The *name* field has two valid values. One value is the name of the element to which the grammar is attached. Currently, this value is for form fields only. The other value is the word "document." Use of the word "document" associates the grammar to a document level context. The name field is optional and, if not specified, the grammar is considered to be a document level grammar.

A document may contain InroadSpeechGrammar references. Multiple references are loaded and attached to the context for their containing document or attached to the context for the appropriate form field.

FIGURE 2 illustrates a portion of an exemplary Web document 150 that is displayed on the display screen 114 (FIGURE 1A) in response to receiving a corresponding HTML document. An HTML code segment for the corresponding HTML document is listed below.

```
<HTML>
```

```
<HEAD>
<TITLE> Bridge HTML (symantec 1.1)</TITLE>
</HEAD>
<BODY>
<!–InroadSpeechGrammar
    grammar = http://www.inroad.site/test.std dictionary =
    http://www.inroad.site/inroad.phd–>
<!–InroadSpeechGrammar
    grammar = http://www.inroad.site/test2.std dictionary =
    http://www.inroad.site/inroad.phd–>
<!–InroadSpeechGrammar name = flavor
    grammar = http://www.inroad.site/dropdn.std dictionary =
    http://www.inroad.site/inroad.phd–>
<!–InroadSpeechGrammar name = age
    grammar = http://www.inroad.site/dropage.std dictionary =
    http://www.inroad.site/inroad.phd–>
        <P> This page is a test page to load grammars for the inroad
        browser.
        </P>
        <a href =
    "http://espnet.sportszone.com">SPORTSZONE</a></p>
        <a href = "http://www.unitedmedia.com/comics/dilbert"
        >DILBERT ZONE</a></p>
        <a href="http://www.usatoday.com">USA TODAY</a>

<SELECT NAME="flavor">
<OPTION VALUE=a>Vanilla
<OPTION VALUE=b>Strawberry
>OPTION VALUE=c>Rum and Raisin
<OPTION VALUE=d>Peach and Orange
</SELECT>

<!–InroadSpeechGrammar name = first
    grammar = http://www.inroad.site/first.std dictionary =
    http://www.inroad.site/inroad.phd–>
```

```
<!–InroadSpeechGrammar name = last
    grammar = http://www.inroad.site/last.std dictionary =
    http://www.inroad.site/inroad.phd–>

<INPUT type = text name=first size=12 maxlength=40>

<INPUT type=text name=last size=12 maxlength=40>

<SELECT NAME="age">
<OPTION VALUE=a>10 to 32
<OPTION VALUE=b>33 to 50
<OPTION VALUE=c>51 to 74
<OPTION VALUE=d>75 to death
</SELECT>

</BODY>
</HTML>
```

In the exemplary HTML code segment listed above, each "grammar=" reference specifies a URL that designates the location of a grammar file. These grammar files are retrieved by the speech-specific parser 206 in the precompiled grammar system 302 of FIGURE 4.

At the Web server, a grammar compiler can be used by a document author to prepare an HTML document for speech recognition. IBM's Via Voice, discussed above, is a grammar compiler that accepts HTML documents and supporting grammar files as input. The toolkit produces speech-enabled HTML documents, grammar files, and dictionary files.

FIGURE 5 illustrates a process 502 of dynamically generating and compiling speech-recognition grammars in accordance with the present invention. At step 504, a new HTML document is received and loaded into the HTML parser 204 (shown in FIGURE 3). At step 506, the HTML parser 204 parses the HTML instructions within the newly received HTML document. At step 506, the HTML parser 204 creates an internal representation 205 of the HTML document. The internal representation includes one or more parse tags. The internal representation is then passed to the speech-specific parser 206.

At step 508, the speech-specific parser 206 retrieves a parse tag from the internal representation of the HTML document. At step 510, a determination is made of whether the retrieved parse tag represents a speakable entity which includes an anchor, form field, text area, button, radio button, check box, and so forth. Speakable HTML entities include anchors, image maps, applets, inputs, and select items. If the current parse tag does not represent a speakable entity, processing proceeds to step 516, where a determination is made of whether the current parse tag is the last parse tag of the current HTML document. If the tag is not the last parse tag, processing returns to step 508 to retrieve the next parse tag.

If, at step 510, the speech-specific parser 206 determines that the current parse tag represents a speakable entity, at step 512, a new rule for the dynamic grammar is created, or, at step 514, an existing rule is appended. The rule adheres to the form:

<rulename> = "Goto link number <n>".

The rule is subsequently used for numerical index navigation. This form is the format used for specifying a grammar rule. The set of rules is then compiled using the grammar compiler, described above.

After creating a rule, processing proceeds to step 516 to determine whether the current parse tag is the last parse tag of the HTML document, as described above. If the tag is not the last parse tag, flow control proceeds back to step 508 to retrieve and process the next parse tag. If, at step 516, the speech-specific parser 206 determines that the current parse tag is the last parse tag, processing proceeds to step 518 where the speech grammar compiler 210 compiles the generated rules into a compiled speech grammar. In one embodiment, the generated rules are in the form of ASCII text, and the compiled speech grammar is a machine representation specific to the speech-recognition engine 212. After compiling the rules to create a compiled speech grammar, the process 502 of dynamically compiling a speech-recognition grammar is complete.

FIGURE 6 illustrates a process 601 that is performed on a wearable computer for displaying a Web document and handling a voice command. At step 602, the wearable computer receives a Web document containing one or more hyperlinks. At step 604, an ordered list of hyperlinks within the Web document is determined. At step 606, the Web document is displayed at the wearable computer.

At step 608, a voice command is received from a user. At step 610, a determination is made of whether the voice command is to display an index menu. If

the command is an index menu command, at step 612, a list of hyperlinks and their corresponding index numbers is displayed. Preferably, the list is displayed within a dialog window. Alternatively, the list may be presented as speech over the wearable computer speakers. After displaying the list of hyperlinks, at step 614, a voice command is received. At step 616, a determination is made of the hyperlink corresponding to the voice command. If, at step 610, the voice command is not an index menu command, processing proceeds to step 616 to determine a corresponding hyperlink. Step 616 may include determining whether the text of a hyperlink was spoken, or whether the index number corresponding to a hyperlink was spoken.

After determining that a hyperlink corresponding to the voice command is present, at step 618, the hyperlink is activated. Activation of the hyperlink may include retrieving a new Web document. Alternatively, activation may comprise displaying a different portion of the same Web document.

While the preferred embodiment of the invention has been illustrated and described, it will be appreciated that various changes can be made therein without departing from the spirit and scope of the invention.

The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows:

1. A system for controlling a web browser using a dynamically generated speech grammar, comprising:

a speech specific parser parsing hyperlinks and interactive controls from a script describing a web page for display by the web browser;

a visual speech hint generator adding a visual speech hint to the web page script for each such hyperlink and visual control;

a speech grammar compiler generating a speech grammar for each such hyperlink and visual control;

a voice audio input device;

a grammar based speech recognition engine translating the speech grammar into a compiled speech grammar and determining a speech event from the voice audio input device using the compiled speech grammar; and

a rendering engine executing the speech event on a visual web page rendered from the modified web page script.

2. system according to Claim 1, further comprising:

a graphical user interface controller translating an input graphical user interface control into a user command, the rendering engine performing an action corresponding to the user command.

3. A system according to Claim 1, further comprising:

a speech controller translating the speech event into a user command, the rendering engine performing an action corresponding to the user command.

4. A system according to Claim 3, further comprising:

a speaker device, wherein the speech controller generates audio prompts using the compiled speech grammar for playing to the user over the speaker device.

5. A system according to Claim 3, wherein the user command is a web browser command.

6. A system according to Claim 1, wherein the speech event is a text message for input into a field within the web page.

7. A system according the Claim 1, further comprising:

a precompiled speech grammar, the grammar based speech recognition engine determining a speech event from the voice audio input using the precompiled speech grammar.

8. A system according to Claim 7, further comprising:

a speaker device;

a speech controller generating audio prompts using the precompiled speech grammar for playing to the user over the speaker device.

9. A system according to Claim 7, wherein the precompiled speech grammar includes non-web page specific web browser commands.

10. A method for controlling a web browser using a dynamically generated speech grammar, comprising:

parsing hyperlinks and interactive controls from a script describing a web page for display by the web browser;

adding a visual speech hint to the web page script for each such hyperlink and visual control;

generating a speech grammar for each such hyperlink and visual control;

translating the speech grammar into a compiled speech grammar;

determining a speech event from voice audio input using the compiled speech grammar; and

executing the speech event on a visual web page rendered from the modified web page script.

11. A method according to Claim 10, the operation of executing the speech event further comprises:

receiving a graphical user interface control;

translating the graphical user interface control into a user command; and

performing an action corresponding to the user command.

12. A method according to Claim 10, the operation of executing the speech event further comprises:

translating the speech event into a user command; and

performing an action corresponding to the user command.

13. A method according to Claim 12, further comprising:

generating audio prompts for the user using the compiled speech grammar.

14. A method according to Claim 12, wherein the user command is a web browser command.

15. A method according to Claim 10, wherein the speech event is a text message for input into a field within the web page.

16. A method according the Claim 10, further comprising:

receiving a precompiled speech grammar; and

determining a speech event from the voice audio input using the precompiled speech grammar.

17. A method according the Claim 10, further comprising:

receiving a precompiled speech grammar; and

generating audio prompts for the user using the precompiled speech grammar.

18. A method according to Claim 17, wherein the precompiled speech grammar includes non-web page specific web browser commands.

19. A computer-readable storage medium containing code for controlling a web browser using a dynamically generated speech grammar, the web browser interfaced with a voice audio input device, comprising:

a speech specific parser parsing hyperlinks and interactive controls from a script describing a web page for display by the web browser;

a visual speech hint generator adding a visual speech hint to the web page script for each such hyperlink and visual control;

a speech grammar compiler generating a speech grammar for each such hyperlink and visual control;

a grammar based speech recognition engine translating the speech grammar into a compiled speech grammar and determining a speech event from the voice audio input device using the compiled speech grammar; and

a rendering engine executing the speech event on a visual web page rendered from the modified web page script.

20. A storage medium according to Claim 19, further comprising:

a graphical user interface controller translating an input graphical user interface control into a user command, the rendering engine performing an action corresponding to the user command.

21. A storage medium according to Claim 19, further comprising:

a speech controller translating the speech event into a user command, the rendering engine performing an action corresponding to the user command.

22. A storage medium according to Claim 21, wherein the web browser is further interfaced with a speaker device, further comprising:

the speech controller generating audio prompts using the compiled speech grammar for playing to the user over the speaker device.

23. A storage medium according the Claim 19, further comprising:

a precompiled speech grammar, the grammar based speech recognition engine determining a speech event from the voice audio input using the precompiled speech grammar.

24. A storage medium according to Claim 23, wherein the web browser is further interfaced with a speaker device, further comprising:

a speech controller generating audio prompts using the precompiled speech grammar for playing to the user over the speaker device.

25. A system for controlling a web browser using a static, precompiled speech grammar, comprising:

a speech specific parser parsing hyperlinks and interactive controls from a script describing a web page for display by the web browser and receiving the precompiled speech grammar;

a visual speech hint generator adding a visual speech hint to the web page script for each such hyperlink and visual control;

a voice audio input device;

a grammar based speech recognition engine determining a speech event from voice audio input using the precompiled speech grammar; and

a rendering engine executing the speech event on a visual web page rendered from the modified web page script.

26. A system according to Claim 25, further comprising:

a graphical user interface controller translating an input graphical user interface control into a user command, the rendering engine performing an action corresponding to the user command.

27. A system according to Claim 25, further comprising:

a speech controller translating the speech event into a user command, the rendering engine performing an action corresponding to the user command.

28. A system according to Claim 27, further comprising:

a speaker device, wherein the speech controller generates audio prompts using the compiled speech grammar for playing to the user over the speaker device.

29. A system according to Claim 27, wherein the user command is a web browser command.

30. A system according to Claim 25, wherein the speech event is a text message for input into a field within the web page.

31. A system according to Claim 25, wherein the precompiled speech grammar includes non-web page specific web browser commands.

32. A method for controlling a web browser using a static, precompiled speech grammar, comprising:

parsing hyperlinks and interactive controls from a script describing a web page for display by the web browser;

receiving the precompiled speech grammar;

adding a visual speech hint to the web page script for each such hyperlink and visual control;

determining a speech event from voice audio input using the precompiled speech grammar; and

executing the speech event on a visual web page rendered from the modified web page script.

33. A method according to Claim 32, the operation of executing the speech event further comprises:

receiving a graphical user interface control;

translating the graphical user interface control into a user command; and

performing an action corresponding to the user command.

34. A method according to Claim 32, the operation of executing the speech event further comprises:

translating the speech event into a user command; and

performing an action corresponding to the user command.

35. A method according to Claim 34, further comprising:

generating audio prompts for the user using the compiled speech grammar.

36. A method according to Claim 34, wherein the user command is a web browser command.

37. A method according to Claim 32, wherein the speech event is a text message for input into a field within the web page.

38. A method according to Claim 32, wherein the precompiled speech grammar includes non-web page specific web browser commands.

39. A computer-readable storage medium containing code for controlling a web browser using a static, precompiled speech grammar, the web browser interfaced with a voice audio input device, comprising:

a speech specific parser parsing hyperlinks and interactive controls from a script describing a web page for display by the web browser and receiving the precompiled speech grammar;

a visual speech hint generator adding a visual speech hint to the web page script for each such hyperlink and visual control;

a grammar based speech recognition engine determining a speech event from voice audio input using the precompiled speech grammar; and

a rendering engine executing the speech event on a visual web page rendered from the modified web page script.

40. A storage medium according to Claim 39, further comprising:

a graphical user interface controller translating an input graphical user interface control into a user command, the rendering engine performing an action corresponding to the user command.

41. A storage medium according to Claim 39, further comprising:

a speech controller translating the speech event into a user command, the rendering engine performing an action corresponding to the user command.

42. A storage medium according to Claim 41, wherein the web browser is further interfaced with a speaker device, further comprising:

the speech controller generates audio prompts using the compiled speech grammar for playing to the user over the speaker device.

43. A method of presenting electronic data comprising:

receiving a document having a plurality of hyperlinks contained therein;

determining an ordered list of hyperlinks from the plurality of hyperlinks, each hyperlink in the ordered list having a corresponding index value representative of the corresponding hyperlink's position in the ordered list of hyperlinks;

displaying the document, wherein the display includes a plurality of hyperlink symbols and a plurality of index symbols, each hyperlink symbol representative of a corresponding hyperlink, each index symbol representative of a corresponding hyperlink's index value;

receiving a voice command to activate a hyperlink;

determining an activated hyperlink contained within the document, wherein the activated hyperlink corresponds to the voice command to activate the hyperlink, wherein determining the activated hyperlink includes determining an index symbol matching the voice command and locating a hyperlink based on the index value corresponding to the index symbol; and

retrieving additional data based on the activated hyperlink.

44. The method of Claim 43, wherein each hyperlink symbol comprises text and each index symbol comprises a number.

45. The method of Claim 43, wherein each index symbol is displayed in close proximity to its corresponding hyperlink symbol.

46. The method of Claim 43, wherein determining the activated hyperlink includes:

determining whether the voice command matches a hyperlink symbol; and

if the voice command matches a hyperlink symbol, locating a hyperlink corresponding to the hyperlink symbol and not locating the hyperlink based on the index value corresponding to the index symbol.

47. The method of Claim 43, further comprising:

receiving a command to display an index menu; and

in response to receiving the command to display an index menu, displaying a list of hyperlink symbols and associated index symbols.

48. A method of presenting electronic data comprising:

receiving a document having a plurality of hyperlinks contained therein;

displaying the document, wherein the display includes a plurality of hyperlink symbols, each hyperlink symbol representative of a corresponding hyperlink;

receiving a voice command to display an index menu;

determining an ordered list of hyperlinks from the plurality of hyperlinks each hyperlink in the ordered list having a corresponding index value representative of the corresponding hyperlink's position in the ordered list of hyperlinks;

in response to receiving the command to display an index menu, displaying a list of hyperlink symbols and associated index symbols; and

receiving a second voice command;

determining an activated hyperlink contained within the document, wherein the activated hyperlink corresponds to the second voice command, wherein determining the activated hyperlink includes determining an index symbol matching the second voice command and locating a hyperlink based on the index value corresponding to the index symbol; and

retrieving additional data based on the activated hyperlink.

49. The method of Claim 48, wherein each hyperlink symbol comprises text and each index symbol comprises a number.

50. A method for providing a voice-activated browser, the method comprising:

generating a voice-recognition grammar at a server computer, the voice-recognition grammar corresponding to a plurality of hyperlink commands corresponding to a Web document;

storing the voice-recognition grammar at the server computer;

associating the voice-recognition grammar with the Web document;

transmitting the Web document and the voice-recognition grammar to a wearable computer;

in response to a voice command at the wearable computer, determining whether the voice command matches a hyperlink command corresponding to the voice-recognition grammar; and

if the voice command matches a hyperlink command corresponding to the voice-recognition grammar, invoking the hyperlink command.

51. The method of Claim 50, further comprising:

determining an ordered list of hyperlinks from the hyperlink commands, each hyperlink command in the ordered list having a corresponding index value representative of the corresponding hyperlink command's position in the ordered list of hyperlinks commands;

displaying the document, wherein the display includes a plurality of hyperlink symbols and a plurality of index symbols, each hyperlink symbol representative of a corresponding hyperlink, each index symbol representative of a corresponding hyperlink's index value; and

wherein determining whether the voice command matches the hyperlink command includes determining whether the voice command matches an index symbol corresponding to the hyperlink command.

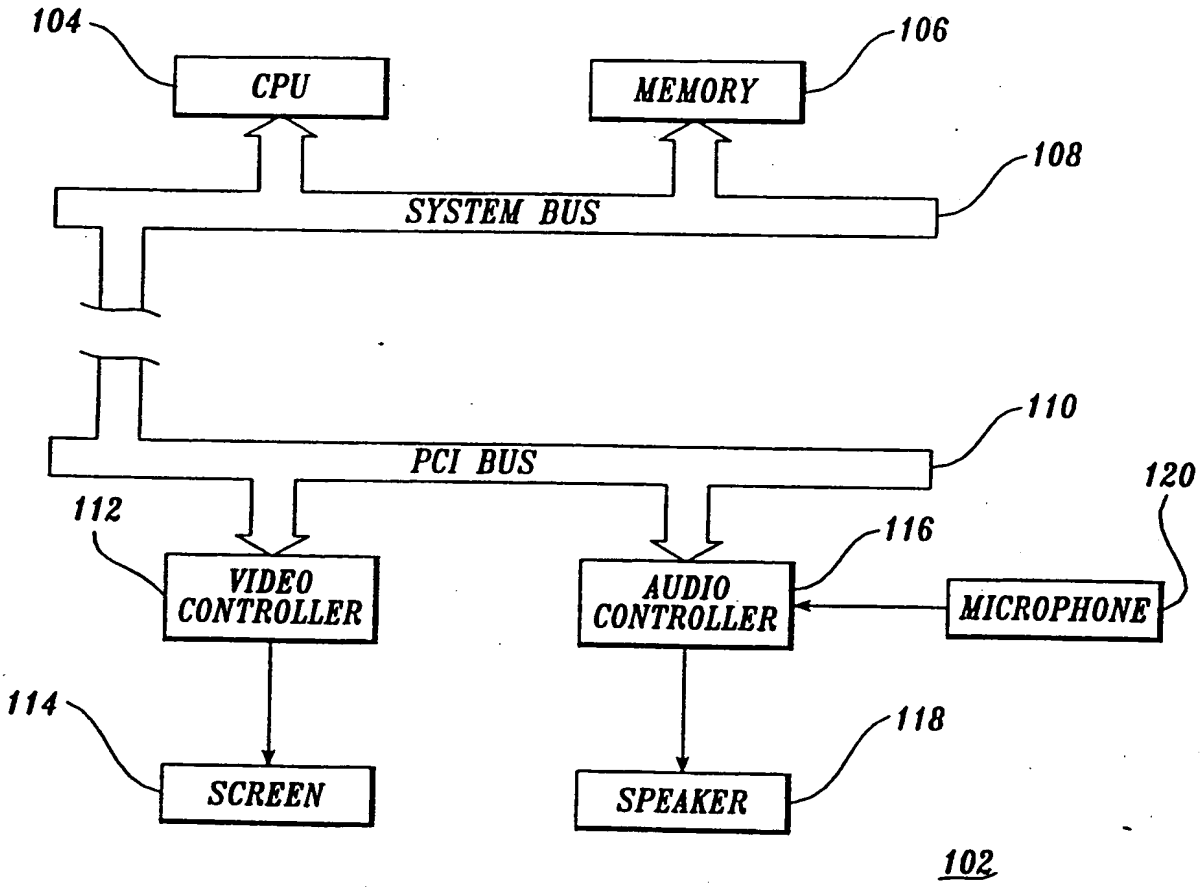
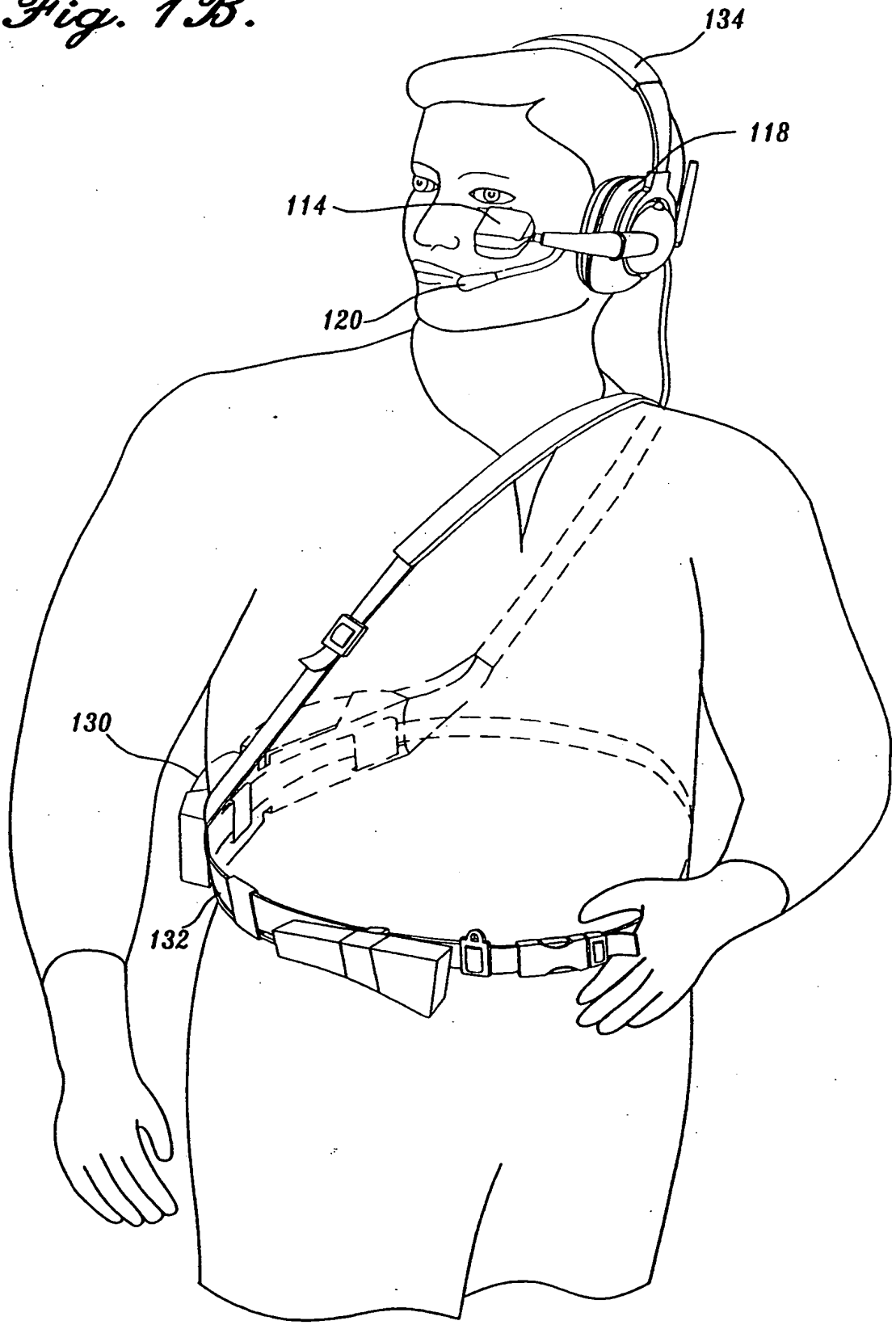

104
CPU
106
MEMORY
108
SYSTEM BUS
110
PCI BUS
120
112
VIDEO CONTROLLER
116
AUDIO CONTROLLER
MICROPHONE
114
SCREEN
118
SPEAKER
102

Fig. 1A.

Fig. 1B.



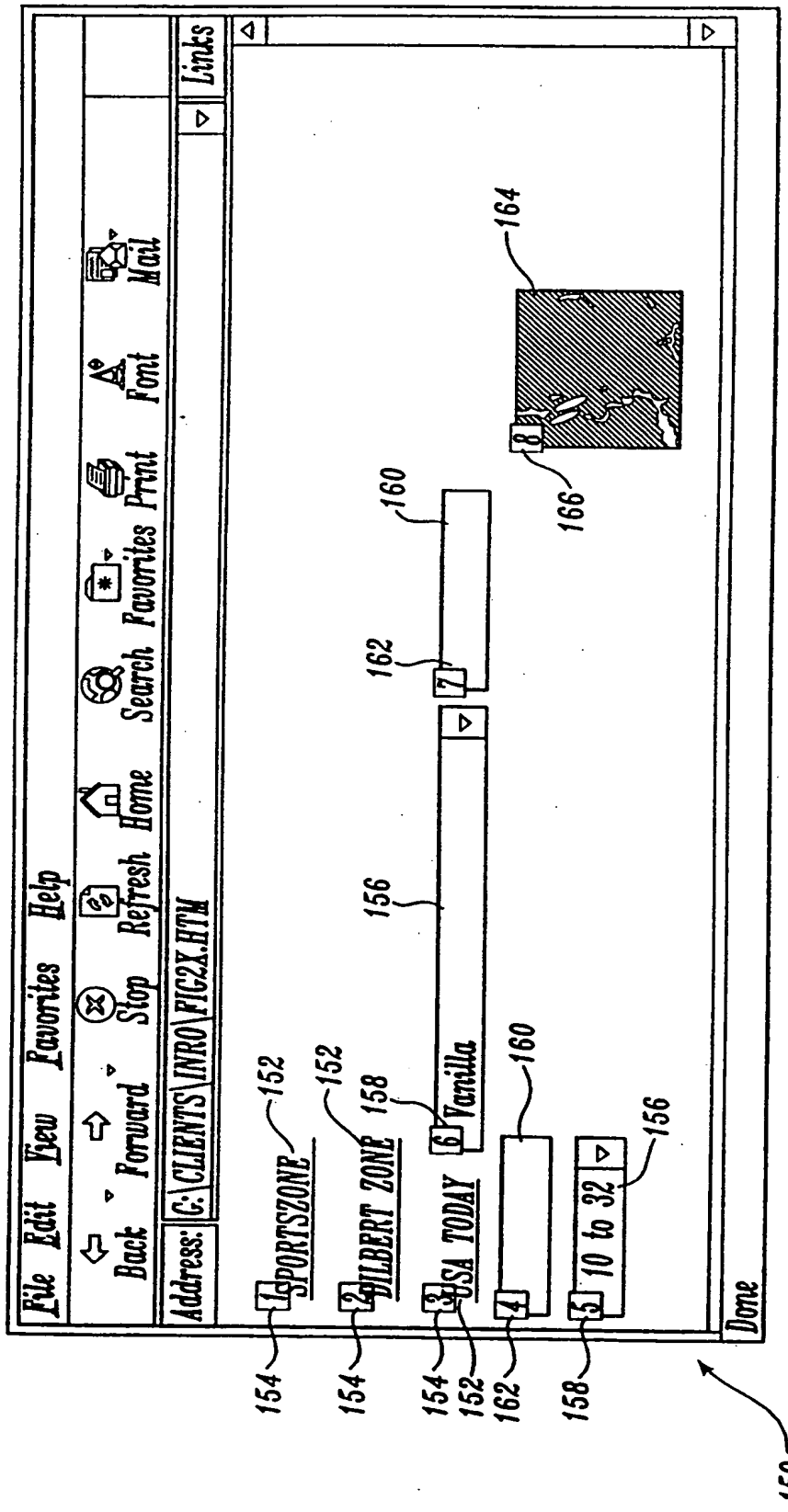

File Edit View Favorites Help
Back Forward Stop Refresh Home Search Favorites Print Font Mail
Address: C:\CLIENTS\INRO\FIG2X.HTM
Links
1 SPORTSZONE
2 DILBERT ZONE
3 USA TODAY
6 Vanilla
7
8
10 to 32
Done
150
152
154
156
158
160
162
164
166

Fig. 2.

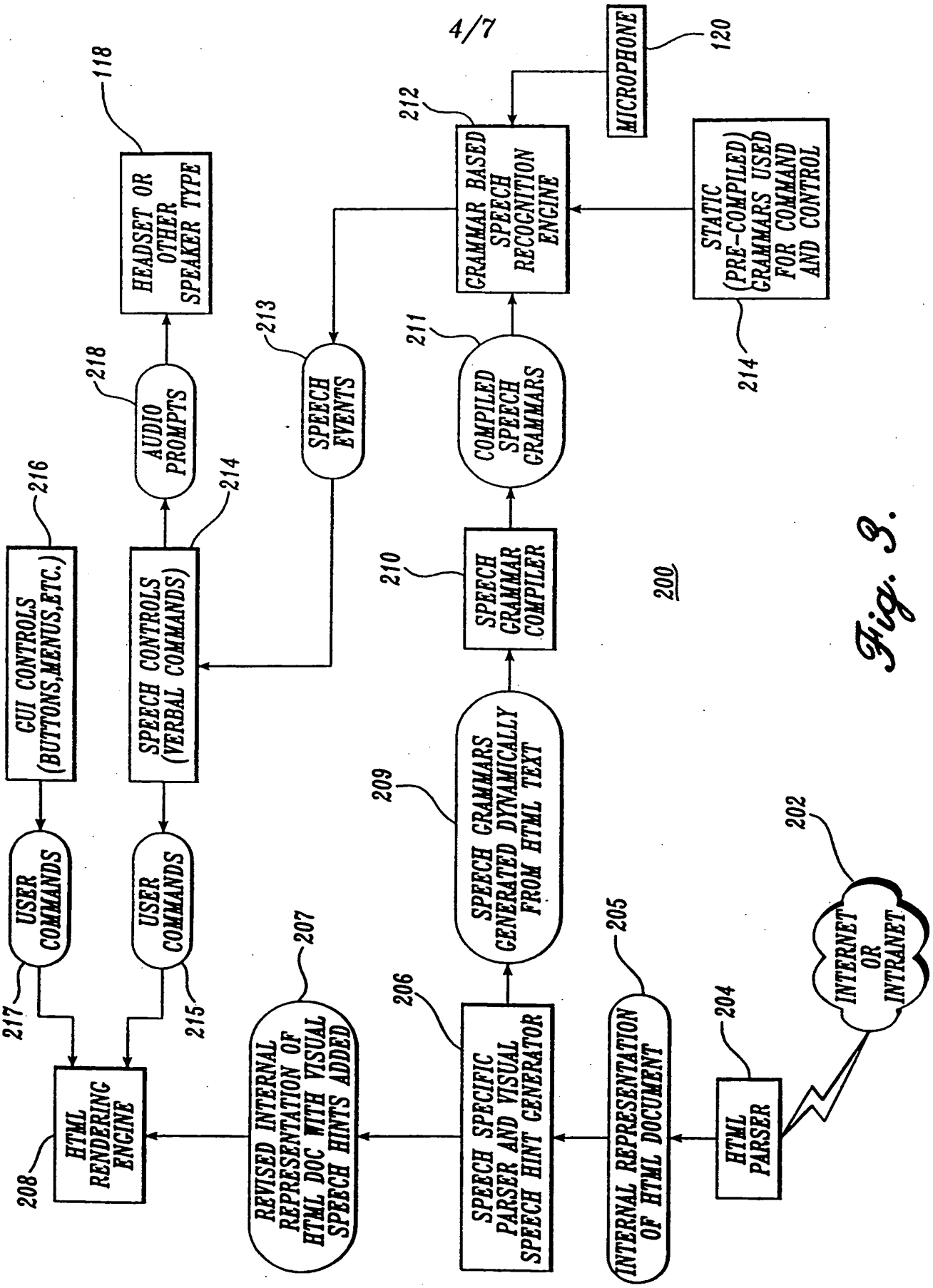

INTERNET OR INTRANET
202
HTML PARSER
204
INTERNAL REPRESENTATION OF HTML DOCUMENT
205
SPEECH SPECIFIC PARSER AND VISUAL SPEECH HINT GENERATOR
206
REVISED INTERNAL REPRESENTATION OF HTML DOC WITH VISUAL SPEECH HINTS ADDED
207
HTML RENDERING ENGINE
208
SPEECH GRAMMARS GENERATED DYNAMICALLY FROM HTML TEXT
209
SPEECH GRAMMAR COMPILER
210
COMPILED SPEECH GRAMMARS
211
GRAMMAR BASED SPEECH RECOGNITION ENGINE
212
MICROPHONE
120
STATIC (PRE-COMPILED) GRAMMARS USED FOR COMMAND AND CONTROL
214
SPEECH EVENTS
213
HEADSET OR OTHER SPEAKER TYPE
118
AUDIO PROMPTS
218
SPEECH CONTROLS (VERBAL COMMANDS)
214
GUI CONTROLS (BUTTONS,MENUS,ETC.)
216
USER COMMANDS
217
USER COMMANDS
215
200

Fig. 3.

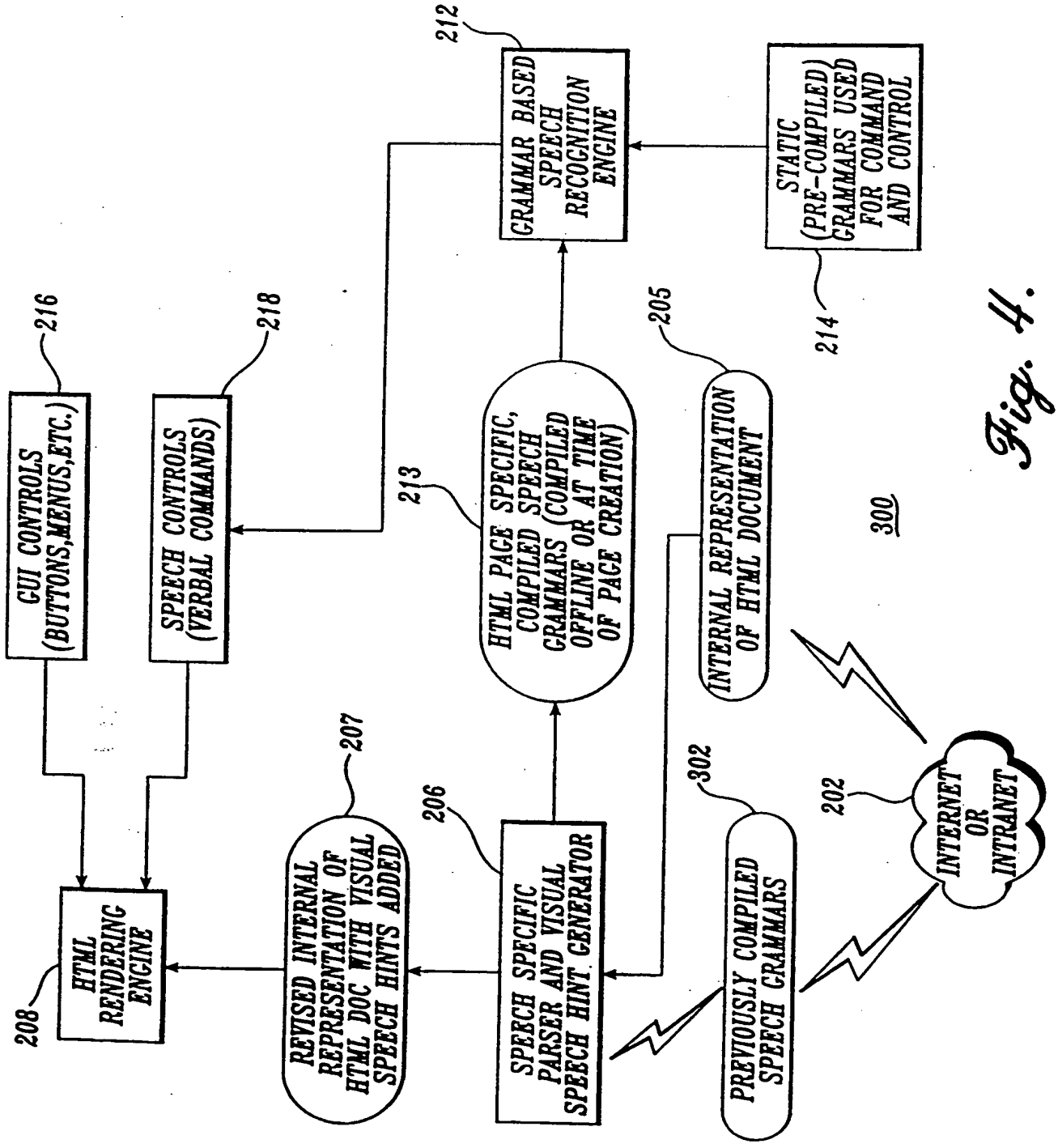

GUI CONTROLS (BUTTONS,MENUS,ETC.)
216
SPEECH CONTROLS (VERBAL COMMANDS)
218
GRAMMAR BASED SPEECH RECOGNITION ENGINE
212
STATIC (PRE-COMPILED) GRAMMARS USED FOR COMMAND AND CONTROL
214
HTML PAGE SPECIFIC, COMPILED SPEECH GRAMMARS (COMPILED OFFLINE OR AT TIME OF PAGE CREATION)
213
INTERNAL REPRESENTATION OF HTML DOCUMENT
205
300
HTML RENDERING ENGINE
208
REVISED INTERNAL REPRESENTATION OF HTML DOC WITH VISUAL SPEECH HINTS ADDED
207
SPEECH SPECIFIC PARSER AND VISUAL SPEECH HINT GENERATOR
206
PREVIOUSLY COMPILED SPEECH GRAMMARS
302
INTERNET OR INTRANET
202

Fig. 4.

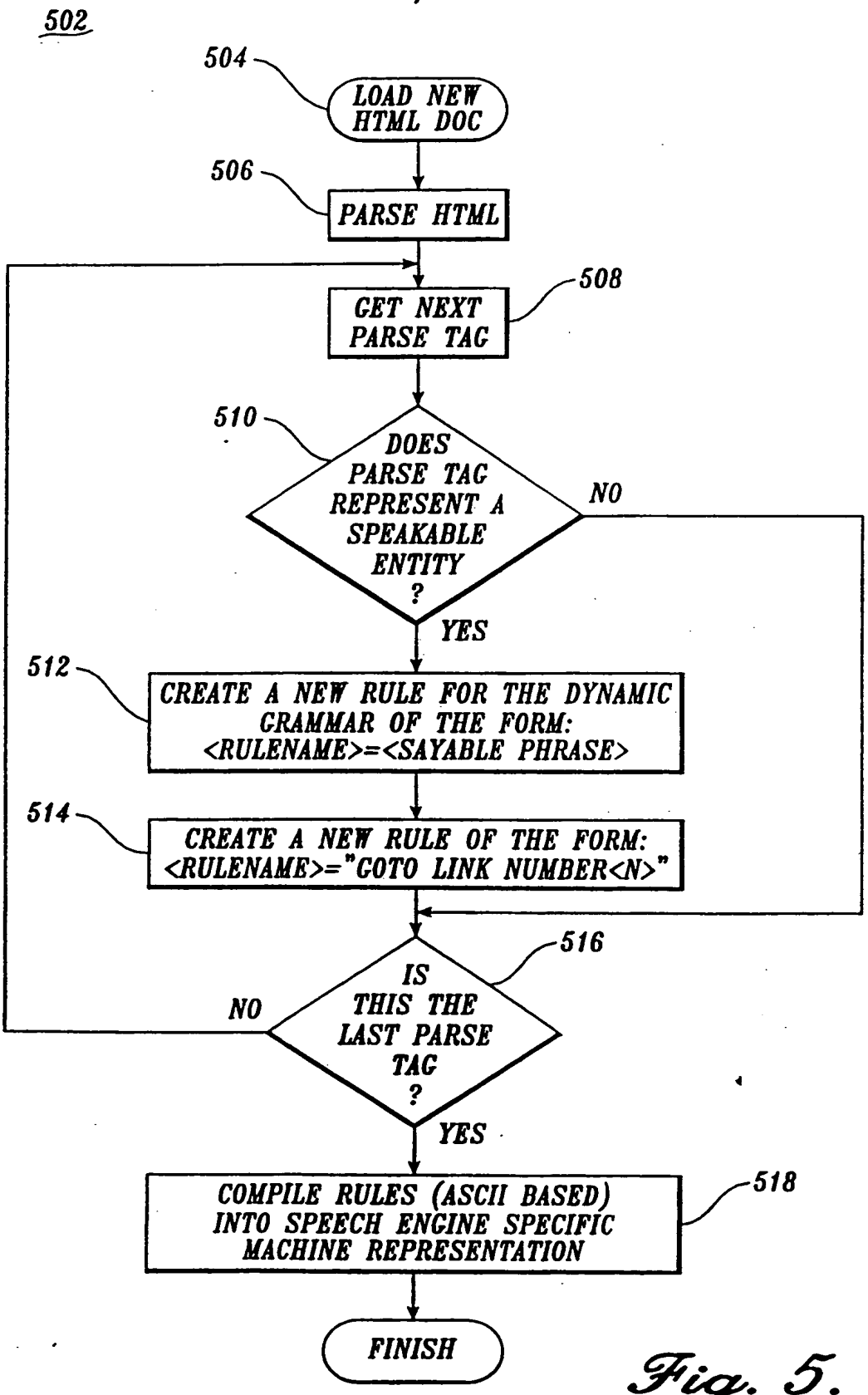

502
504
LOAD NEW HTML DOC
506
PARSE HTML
508
GET NEXT PARSE TAG
510
DOES PARSE TAG REPRESENT A SPEAKABLE ENTITY ?
NO
YES
512
CREATE A NEW RULE FOR THE DYNAMIC GRAMMAR OF THE FORM: <RULENAME>=<SAYABLE PHRASE>
514
CREATE A NEW RULE OF THE FORM: <RULENAME>="GOTO LINK NUMBER<N>"
516
IS THIS THE LAST PARSE TAG ?
NO
YES
518
COMPILE RULES (ASCII BASED) INTO SPEECH ENGINE SPECIFIC MACHINE REPRESENTATION
FINISH

Fig. 5.

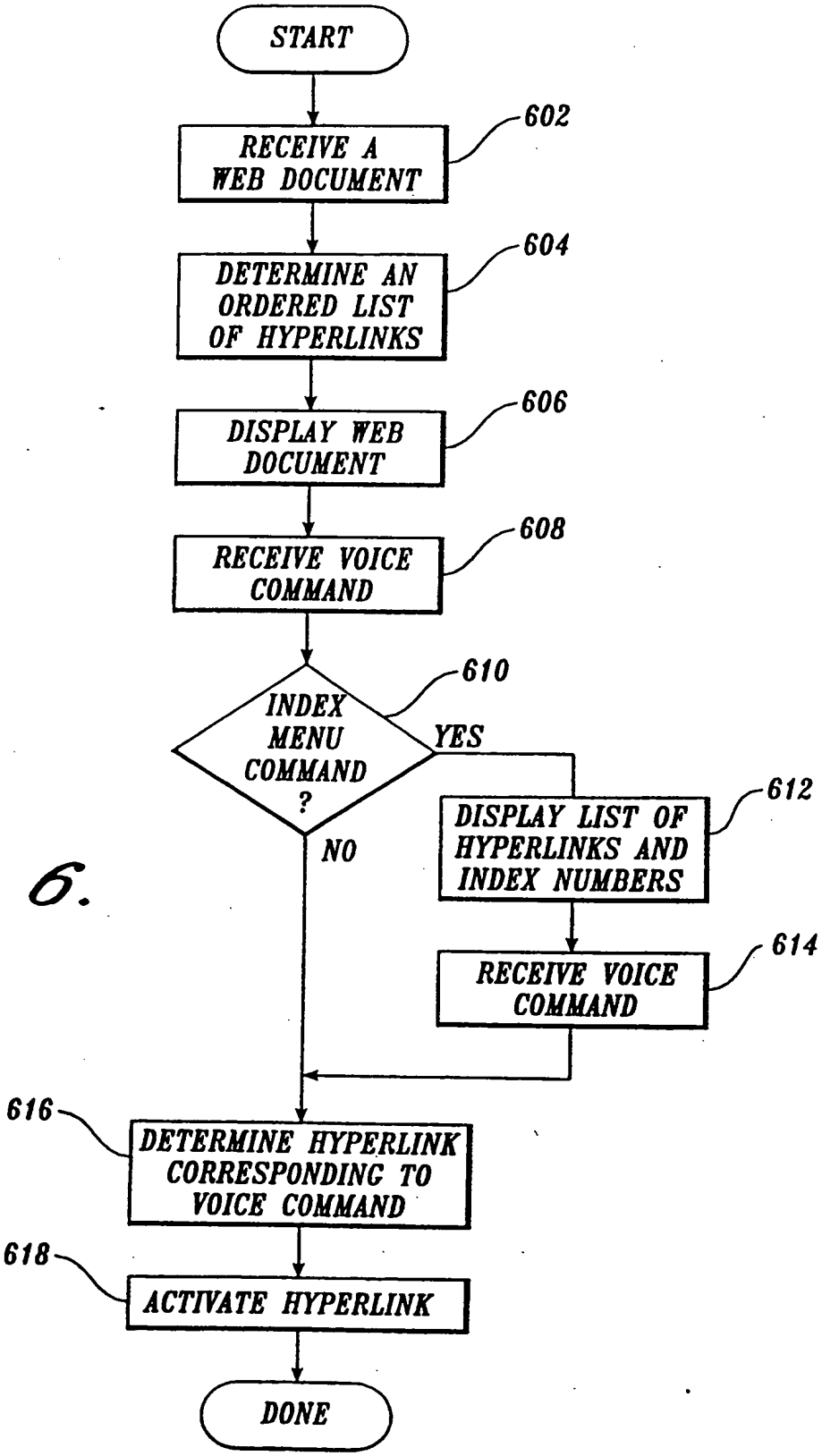

START
RECEIVE A WEB DOCUMENT
602
DETERMINE AN ORDERED LIST OF HYPERLINKS
604
DISPLAY WEB DOCUMENT
606
RECEIVE VOICE COMMAND
608
INDEX MENU COMMAND ?
610
YES
NO
DISPLAY LIST OF HYPERLINKS AND INDEX NUMBERS
612
RECEIVE VOICE COMMAND
614
DETERMINE HYPERLINK CORRESPONDING TO VOICE COMMAND
616
ACTIVATE HYPERLINK
618
DONE

Fig. 6.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/06072

A. CLASSIFICATION OF SUBJECT MATTER
IPC(6) :G10L 9/06
US CL :704/275
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 704/275

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, IEL
search terms: Web browser, speech, hyperlink

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A, P	US 5,819,220 A (SARUKKAI et al) 06 October 1998, see abstract	1-51
A	HEMPHILL et al. Speech-Aware Multimedia. IEEE Multimedia. Spring 1996. Col. 1, Issue 1.	1-51
A	ZUE, V.W. Navigating the Information Superhighway using Spoken Language Interfaces. IEEE Expert. October 1995. Vol. 10, Issue 5. pages 39-43	1-51
A	BAYER, S. Embedding Speech in Web Interfaces. ICSLP 96. Proceedings., Fourth International Conference on Spoken Language, 1996. October 1996. Vol. 3	1-51

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 01 JUNE 1999	Date of mailing of the international search report 14 JUN 1999
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer M. DAVID SOPOCLEOUS Telephone No. (703) 308-4825

Form PCT/ISA/210 (second sheet)(July 1992) *

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/06072

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A, P	LAZZRO, J.J. Helping the Web Help the Disabled. IEEE Spectrum. March 1999. Vol. 36, Issue 3.	1-51
A	KANEEN et al. A Spoken Language Interface to Interactive Multimedia Services. IEEE Colloquium on Adavnces in Interactive Voice Technologies for Telecommunication Services. June 1997.	1-51